

Religious Fundamentalism and Far-Right Ideology in Brazilian WhatsApp and Telegram Groups

A Multidimensional Computational Analysis

Fernanda Nascimento
Universidade Federal do Ceará
(UFC)
Fortaleza, Ceará, Brazil
fernanda.nascimento@alu.ufc.br

Melissa Sousa Felipe
Universidade Federal do Ceará
(UFC)
Fortaleza, Ceará, Brazil
melissasousa@alu.ufc.br

Letícia Torres
Universidade Federal do Ceará
(UFC)
Fortaleza, Ceará, Brazil
leticia.torres@alu.ufc.br

Ivandro Claudino
Universidade Federal do Ceará
(UFC)
Fortaleza, Ceará, Brazil
ivandro.claudino@gmail.com

Jose Maria Monteiro
Universidade Federal do Ceará
(UFC)
Fortaleza, Ceará, Brazil
monteiro@dc.ufc.br

Javam Machado
Universidade Federal do Ceará
(UFC)
Fortaleza, Ceará, Brazil
javam.machado@dc.ufc.br

ABSTRACT

This paper examines the relationship between religious fundamentalism and far-right political ideology in Brazilian public groups on WhatsApp and Telegram through a multidimensional computational analysis. To this end, we constructed two datasets comprising religious messages circulated in far-right public groups on both platforms. We then applied Named Entity Recognition to identify the most frequent entities, which were subsequently used to retrieve messages for manual discourse analysis. Each message was then assigned a sentiment polarity label. We also applied three topic modeling algorithms (LDA, PTM, and BERTopic) to identify latent thematic structures and used LIWC to examine emotional, cognitive, and linguistic dimensions. The results reveal significant psycholinguistic differences between the two platforms, despite the emergence of similar thematic patterns across the three topic modeling approaches. Overall, this study provides empirical evidence of how religious fundamentalist discourse circulates and becomes politically articulated within Brazilian far-right groups on WhatsApp and Telegram.

KEYWORDS

religious fundamentalism, far-right, whatsapp, telegram

1 INTRODUCTION

In recent years, we have witnessed a remarkable growth of Christian fundamentalist groups in Brazil, which operate and organize primarily through social media [4, 6, 36]. Christian fundamentalism can be understood as the pursuit and desire to preserve the foundations of Christianity [5] in the face of the advancement of modernity. One of the main characteristics of Christian fundamentalism is a literal interpretation

of sacred texts. From this type of decontextualized reading emerge further traits such as intolerance, xenophobia, racism, machismo, sexism, and homophobia [3, 4, 12, 17, 27]. It is worth noting that there is an affinity between far-right movements and fundamentalist groups, since both seek to maintain a certain natural order of things [14, 40]. The far-right is a political ideology whose main ideals are nationalism, xenophobia, racism, misogyny, and anticommunism [35, 38].

Christian fundamentalism carries an interesting contradiction. It emerges as a reaction to modernism, as it does not tolerate scientific advances, yet it extensively utilizes contemporary communication media, including instant messaging platforms such as WhatsApp and Telegram [17]. WhatsApp is the most widely used social network in Brazil, available on practically all smartphones in the country [19]. On the other hand, in just one year, the proportion of *smartphones* with Telegram in Brazil grew from 45% to 60% in 2022 [18].

In this context, it is of paramount importance to characterize the dissemination of religious fundamentalist content on instant messaging platforms and its convergence with far-right political ideologies in Brazilian society. However, there are few comprehensive empirical studies characterizing fundamentalist religious discourse in public WhatsApp and Telegram groups in Brazil, despite growing concerns about Christian religious fundamentalism in the country. This paper aims to address the following research question:

- RQ1: How is the relationship between religious fundamentalism and far-right political ideology in messages shared on WhatsApp and Telegram?
- RQ2: How does fundamentalist discourse manifest emotionally across WhatsApp and Telegram platforms?
- RQ3: What are the predominant latent topics in Brazilian religious fundamentalist discourse propagated on WhatsApp and Telegram?
- RQ4: Are there psycholinguistic differences between messages shared on WhatsApp and Telegram?

In: Proceedings of the Brazilian Symposium on Multimedia and the Web (WebMedia'2026). Lavras/MG, Brazil. Porto Alegre: Brazilian Computer Society, 2026.

© 2026 SBC – Brazilian Computing Society.
ISSN 2966-2753

To answer these questions, we constructed two datasets containing religious messages circulated in far-right public groups on both platforms. Next, we applied Named Entity Recognition (NER) to identify the most frequent entities in the texts collected from each platform. These entities were subsequently used as search terms in a text retrieval system to locate the messages in which they appeared. The retrieved messages were then manually examined to provide a more detailed understanding of the discourse present in each dataset. We also assigned a sentiment polarity label to each message. To investigate the underlying thematic structure of the datasets, we applied three different topic modeling approaches. Finally, the LIWC framework was employed to analyze the emotional, cognitive, and linguistic characteristics of the texts. The findings indicate that religious fundamentalist contents are present in far-right public groups (RQ1). The two platforms seems quite equivalent in emotional polarity since messages on Telegram are only marginally more negative than those on WhatsApp (RQ2). Topic analysis revealed distinct topics for each examined platform (RQ3). The LIWC-based linguistic analysis identified significant differences in 49 of the 63 psycholinguistic categories, but with uniformly negligible effect sizes (RQ4).

The structure of this paper is organized as follows: Section 2 discusses related works, Section 3 describes the methodological design adopted to address the research questions, and Section 4 presents the results of several analyses. Finally, limitations and conclusions are presented in Section 5.

2 RELATED WORK

In [10], the authors aim to model and characterize hate speech against immigrants on Twitter in Spain. More than 240,000 tweets posted between November 2018 and April 2019 were collected and automatically analyzed. Using the LDA (Latent Dirichlet Allocation) algorithm[9], four topics were identified: illegal immigration control, economic assistance for immigrants, the consequences of illegal immigration, and Islamic terrorism. The study presented in [41] proposes a methodology that combines LDA with self-organizing maps to analyze hate speech disseminated on social media.

Kursuncu et al. [30] modeled Islamist extremist communications on Twitter by jointly considering three contextual dimensions—religion, ideology, and hate—to support the analysis of online radicalization. The study presented in [26] explores how Twitter users conceptualize extremism in Arabic and articulate their arguments to construct the term. A corpus of tweets containing the word “extremism” was collected using the Twitter API. The LDA algorithm was then applied to the dataset to uncover latent concepts associated with extremism. In [1], a sophisticated approach is proposed to analyze extreme views expressed on social media. The proposed approach combines machine-learning algorithms, natural language processing techniques, and expert-driven content analysis. A dataset comprising more than 259,000 tweets was used to evaluate the approach, yielding several insights into far-right extremism.

In [25], the authors examined the main publicly available datasets for hate speech research, aiming to identify the types of hate speech (e.g., ethnicity, religion, sexual orientation) represented in their composition. The results provided *insights* that contribute to a better understanding of the phenomenon of hate speech and to improving its detection on social platforms. In [44], the authors present a system called “NoHateBrazil” that recognizes both explicit and implicit offenses. The proposed system encompasses data collection, human annotation, and the training of machine-learning models, which were used to automatically detect the presence of hate speech.

In [43], the authors conducted a comparative analysis of the most frequently used Natural Language Processing (NLP) techniques, discussing the availability of datasets, and highlighting the *insights* reported by the reviewed studies. In [23], the authors present a systematic review of the analysis and detection of content related to extremism, radicalization, and hate speech. A total of 51 articles published between 2015 and 2021 were analyzed. Finally, the paper outlines a set of recommendations for researchers, as well as guidelines for a safer cyberspace.

Cavalini et al. [15] characterized over 2 million messages from 128 Brazilian extremist Telegram chats, showing that content toxicity correlates with its spread across the network. Piorino et al. [37] evaluated how toxic content operates as a “trigger” for extending and inflaming discussions within Brazilian Reddit communities, while mapping lexicons and frequent terms with negative or polarizing connotations.

Lobo et al. [32] investigated the dynamics of hate speech propagation through toxicity cascades in Twitch live chats using NLP tools including lexical dictionaries and deep learning models, mapping 472 cascades and identifying patterns of collective toxicity. Gibrin et al. [22] presented a systematic review of the scientific literature on hate speech in games, covering publications from 2010 to 2025.

Martins et al. [33] examined the presence and activity of the Red Pill community on YouTube Brazil, aiming to characterize how misogynistic discourse is constructed, interpreted, and propagated. They analyzed 18,034 videos and over 2.2 million comments from 28 channels published between January 2024 and June 2025, applying natural language processing techniques to investigate engagement patterns and discursive dynamics.

Our work differs from the previous studies along three main dimensions. First, in terms of phenomenon, we focus specifically on the articulation between Christian religious fundamentalism and far-right ideology in the Brazilian context, rather than on extremism, radicalization, or hate speech. Second, in terms of platforms, we investigate public groups from both WhatsApp and Telegram, instant-messaging environments that remain underexplored in computational studies of political and religious extremism. Third, in terms of analytical design, rather than relying on a single computational perspective, we combine different methods in a multidimensional design.

3 MATERIALS AND METHODS

This section presents the methodological design adopted to address the four research questions previously defined, covering the stages of data collection, anonymization, and preprocessing, as well as the application of different analyses aimed at identifying thematic, discursive, and psycholinguistic patterns. Figure 1 illustrates the end-to-end workflow adopted to address the four defined research questions. In order to ensure the study's transparency and reproducibility, all *scripts* used are publicly available in our GitHub repository¹.

3.1 The Explored Datasets

We aim to characterize the relationship between religious fundamentalism and far-right political ideology in Brazilian public groups on WhatsApp and Telegram. Thus, we constructed two datasets, named FW.Br and FT.Br, respectively.

The FW.Br dataset was built from messages collected from public WhatsApp groups associated with the Brazilian far right, using the platform proposed in [19], over the period from September 27, 2022 to November 14, 2022. Initially, we obtained 598,971 messages from 15,286 distinct users across 240 different groups. Since our goal is to characterize the religious discourse that spreads in these groups, not all messages were relevant. To retain only religious messages, we applied an inclusion filter based on a set of keywords. To define the keywords used in this filter, we adopted the following approach. First, we asked three professionals with undergraduate degrees in Theology to provide a list of words associated with Christian religious fundamentalism. Next, we selected the words that appeared in the intersection of the three previously produced lists. This procedure yielded a set of 18 keywords (in Portuguese): God, Jesus, mercy, David, Solomon, kingdom, temple, conservative, pentecostal, rcc, renewal, charismatic, Paulo Ricardo, Bernardo Kuster, heretic, atheist, Jericho, and heresy. We then applied the established filter, selecting messages that contained at least one of the 18 previously defined keywords. It is worth noting that RCC stands for Catholic Charismatic Renewal (Renovação Carismática Católica, in Portuguese) and that Paulo Ricardo and Bernardo Küster are highly popular Catholic digital influencers in Brazil whose content is commonly associated with fundamentalist orientations. Subsequently, due to the strong presence of content related to election campaigns, we applied an exclusion filter based on keywords (in Portuguese), removing messages that contained any of the following terms: lula, bolsonaro, pt, pl, stf, pátria, 55, 22, 13, senadores, luladrão, urnas, alexandre, moraes, comunismo, eleição, eleições, esquerda, direita, presidente, tse, fraude, voto, turno, and ministro. The presence of terms associated with election campaigns introduces noise into the dataset and hinders data analysis. This set of keywords was found empirically. After applying the inclusion and exclusion filters, the FW.Br dataset was reduced from 598,971 messages to 5,167.

Similarly, the FT.Br dataset consists of messages collected from public Telegram groups. Initially, 557,586 messages were obtained from 14,810 distinct users across 178 different groups. Next, we applied the same strategy used for WhatsApp. After applying the inclusion and exclusion filters, the number of messages dropped to 8,526. Table 1 presents the basic statistics computed for the FW.Br and FT.Br datasets, including the number of unique messages, the minimum and maximum number of *tokens*, as well as the mean and standard deviation of the *token* count. From Table 1, it can be seen that the characteristics of the messages in FW.Br and FT.Br do not appear to change substantially.

Figure 2 shows the number of messages collected per day for each of the platforms used, following the application of filters. The temporal analysis of message volume reveals striking differences between the platforms. While WhatsApp shows a more stable activity pattern with moderate peaks, Telegram exhibits significantly higher message volumes, accompanied by abrupt peaks in engagement, especially during politically sensitive periods. Furthermore, we can observe a peak from the end of October to the beginning of November, both on Telegram (with almost 600 messages in a single day) and on WhatsApp (with almost 300 messages), which corresponds to the period of the Brazilian general elections. This finding indicates that the number of messages containing terms related to religious fundamentalism shared in far-right groups increased during periods of heightened political tension.

3.2 Data Protection and Pseudonymization

To protect users' privacy and comply with Brazil's General Data Protection Law (LGPD), identifying attributes such as telephone numbers and group identifiers were pseudonymized through cryptographic hash functions before the analytical stages. The study did not seek to identify individual users, and these identifiers were used only when necessary to preserve structural information relevant to the analyses. The reported results focus primarily on aggregate patterns and do not intentionally disclose identifying information about non-public individuals.

3.3 Data Analysis Steps

In this study, we investigate the relationship between religious fundamentalism and far-right political ideology in Brazilian public WhatsApp and Telegram groups. To do this, we will explore different data analysis techniques. All datasets and experiments performed are available online².

3.3.1 Named Entities Recognition (NER). Named entities are textual expressions that represent specific and well-defined elements of the real world, usually with proper names. Named Entity Recognition (NER) is a task that consists of automatically detecting named entities in a text, as well as classifying them into predefined categories (people, organizations, places, monetary values, dates, and times). In this work, the

¹<https://github.com/jmmfilho/fundamentalismoreligioso-whatsapp-telegram>

²<https://anonymous.4open.science/r/fundamentalismoreligioso-whatsapp-telegram-7F37/>

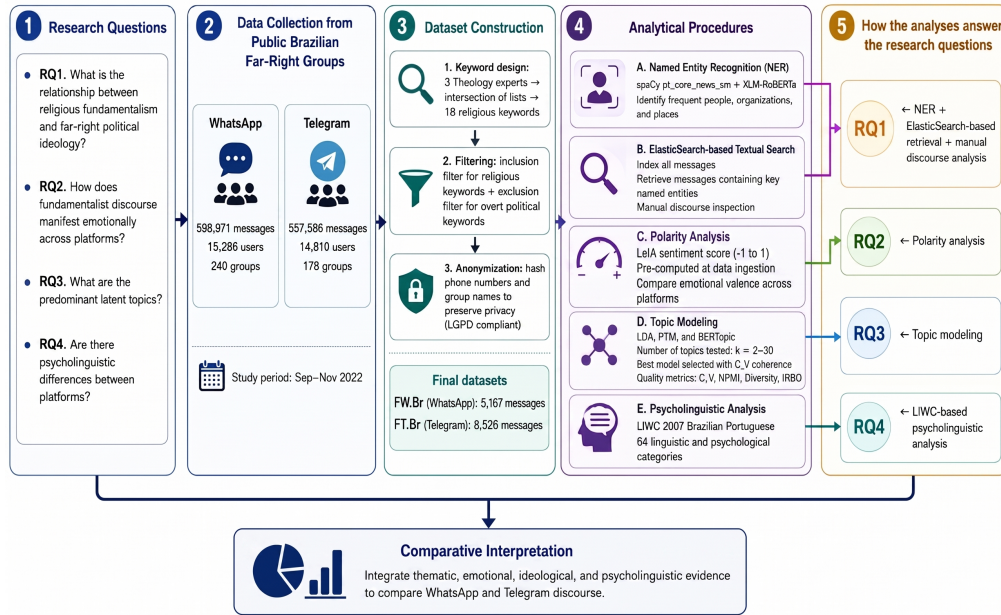


Figure 1: Overview of the Research Methodology. The end-to-end workflow adopted to address the four research questions.

Table 1: Basic Statistics of FW.Br and FT.Br Datasets

Statistics	FW.Br (WhatsApp)	FT.Br (Telegram)
Number of unique messages	4.329	6.462
Mean and standard-deviation of token numbers	82,34 ± 99,64	88,57 ± 116,87
Minimum number of tokens	3	3
Maximum of tokens	1.683	1.549

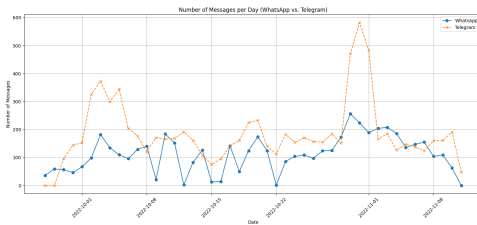


Figure 2: Per Day Message Distribution

spaCy library, along with the pre-trained Portuguese model *pt_core_news_sm* [21], and the XLM-RoBERTa model [16] were used to recognize named entities. The most frequently named entities in the text collected from a given platform can provide *insights* into the preferences and ideological positions of its users.

3.3.2 Textual Search with ElasticSearch. A text search engine is a computational system that indexes, retrieves, and ranks textual documents to efficiently answer queries. Elasticsearch is a distributed search engine designed for real-time analysis

of large data volumes [29]. All collected messages were indexed using Elasticsearch, enabling the previously identified named entities to be used as search terms to retrieve the messages in which they appeared. The retrieved messages were then manually examined to gain insights into the discourse represented in each dataset.

3.3.3 Polarity Analysis. Polarity analysis is a NLP technique that aims to automatically identify the emotional valence expressed in a text by assigning it a polarity category (positive, negative, or neutral) based on lexical, semantic, or contextual content. In this work, polarity was computed using the LeIA library³, a fork of the VADER lexicon adapted to Portuguese, with emoji support and a focus on social media text [2, 28]. LeIA assigns a value normalized to $-1, 1$ to each message. Messages scoring above 0.05 are classified as positive, below -0.05 as negative, and the intermediate range as neutral.

3.3.4 Topic Modeling. Topic modeling is an unsupervised method used to identify latent themes within large text collections. In this study, it is employed to identify, characterize,

³<https://github.com/rafjaa/LeIA>

and compare the main thematic dimensions of religious discourse in far-right groups on WhatsApp and Telegram, while accounting for the discursive specificities of each platform.

To capture distinct semantic dimensions of the corpus, we employed topic models with different methodological foundations. LDA (Latent Dirichlet Allocation) [9], a probabilistic *bag-of-words*-based model, was used to identify broad and recurrent themes, providing high interpretability and an overall view of the corpus's thematic structure. PTM (Pseudo-document-based Topic Model) [45] extends LDA by grouping short messages into pseudo-documents, thereby reducing the sparsity and noise commonly associated with messaging data and making it particularly suitable for WhatsApp and Telegram. BERTopic [24], which combines contextual *embeddings* with clustering techniques, captures subtler and context-dependent semantic relationships, enabling the identification of more fine-grained topics, narrative variations, and discursive framings that may not emerge from lexical co-occurrence alone.

The combined use of these models enables a complementary and comparative analysis of the identified topics, improving the robustness of the results while mitigating biases associated with individual modeling assumptions. The number of topics (k) was varied from 2 to 30, and the optimal configuration was selected using the C_V coherence metric. Topic quality was further evaluated using C_V, NPMI (*Normalized Pointwise Mutual Information*), diversity, and IRBO (*Inverted Rank-Biased Overlap*), as implemented in the OCTIS library [42].

3.3.5 Linguistic Analysis. The LIWC (*Linguistic Inquiry and Word Count*) tool is a dictionary-based software that analyzes texts and provides statistics on the frequency of words belonging to psychological, social, and linguistic categories, enabling inferences about the author's emotional state, language style, and cognitive traits [13]. In this work, the psycholinguistic analysis of the transcribed texts was performed using the LIWC 2007 dictionary adapted to Brazilian Portuguese [7], which includes 64 categories distributed across psychological, linguistic, and social dimensions. LIWC computes the percentage of words in each category relative to the total number of words in the text, allowing inferences about emotional, cognitive, and structural aspects of discourse. Examples of categories used include: *positive emotions* (e.g., “love”, “happiness”), *negative emotions* (e.g., “anger”, “fear”), *cognitive processes* (e.g., “think”, “because”, “understand”), *social words* (e.g., “family”, “friend”), and *personal pronouns* (e.g., “I”, “we”).

4 DATA ANALYSIS

This section presents the main results obtained through the application of different data analysis techniques on religious messages shared on WhatsApp and Telegram far-right groups. The analysis of named entities, polarity, topics, and psycholinguistic aspects is described.

4.1 Named Entity Recognition

As mentioned previously, named entities are expressions that play a key role in understanding a given context and Named Entity Recognition (NER) is the process that aims to identify these entities in a textual corpus and classify them into predefined categories. Table 2 presents the 10 most frequent named entities in each analyzed dataset. Notably, 4 of the 10 entities (40%) occur in both datasets. A strong presence of geopolitical entities can also be observed, including Brazil, the United Kingdom, the USA, and Venezuela. This finding reinforces the close relationship between religious fundamentalism and far-right ideology. In this analysis, we chose to retain both unidentified proper names (e.g., Chirley and Edson) and the names of public figures (e.g., Hostílio Souza).

Table 2: Top Named Entities

spaCy	XLM-RoBERTa
FW.Br	
<i>Jesus, Brazil, Jesus Christ, God, iFood, Chirley, Hostílio Souza, mercy, belief, family</i>	<i>Brazil, je, Jesus, sus, Israel, Ley, Galilee, Jesus Christ, Reginaldo Manzotti</i>
FT.Br	
<i>Brazil, Jesus, Jesus Christ, United Kingdom, God, Jesus saves, read the Bible, USA, Brazil, Venezuela, Edson</i>	<i>Brazil, Jesus, je, sus, da, Brazil, Israel, Mari, God</i>

4.2 Textual Search with ElasticSearch

Based on the previously identified named entities, we issued queries using a textual search engine, specifically the tool called *ElasticSearch*⁴, in order to retrieve relevant messages. Next, each retrieved message was manually read and analyzed by a specialist in Theology.

Regarding the FW.Br dataset, the most frequently shared message (53 times) is an “*exposed*”⁵ containing the name and business address of a company owned by a woman who posted a video criticizing the political instrumentalization of evangelical faith. Concerning the FT.Br dataset, the most shared message (48 times) is a video featuring Dercy Gonçalves commenting on the popularity of Jesus Christ. It is also worth highlighting a schismatic message (shared eight times) that challenges the authority and teaching office of Pope Francis. Upon analyzing the selected messages, it becomes evident that Brazilian religious fundamentalism is closely linked to far-right political movements, and that spiritual warfare has tangible repercussions in the material world, posing concrete threats to individuals, groups, and institutions. Additionally, we observe a strong connection between religious and political content linked to the far right, including explicit voting guidance and allegations of fraud in the electoral system. Comparing the two datasets, FW.Br and FT.Br, we can note that the messages obtained from Telegram are somewhat more radical than the texts collected from WhatsApp.

⁴<https://www.elastic.co/pt/elasticsearch>

⁵An English term meaning to publicly reveal or expose information about a fact, person, or company

4.3 Polarity Analysis

Now, we will analyze the shared messages based on the “*Sentiment Score*” metric, which expresses the intensity and emotional direction of the text. This metric assumes continuous values between -1 and 1, in which values close to -1 indicate negative sentiments, values close to 1 indicate positive sentiments, and values close to 0 reflect neutrality. Table 3 reports the descriptive statistics of the *Sentiment Score* in both datasets. The two distributions are bimodal and predominantly positive: 64.8% of the messages in FW.Br and 60.8% in FT.Br fall in the positive range, against 25.4% and 28.4% in the negative range. This positive valence reflects the devotional register of the retrieved messages — blessings, invocations, calls to mobilization — whose vocabulary carries positive lexical valence even when the message conveys hostility. Since the score is bounded and markedly non-normal, we compared the platforms using the Mann-Whitney U test, reporting Cliff’s δ as effect size. FT.Br messages are slightly more negative, and the difference is significant ($U = 22,820,276$, $p = 4.0 \times 10^{-4}$), but the effect size is negligible ($\delta = 0.036$, 95% CI 0.016, 0.056); the proportion of negative messages is only 3.0 percentage points higher in FT.Br ($\chi^2 = 21.28$, Cramér’s $V = 0.039$)⁶. We therefore conclude that the two platforms are equivalent in lexical emotional polarity, in contrast with the psycholinguistic differences reported in Section 4.5.

Table 3: Basic Statistics About Polarity Analysis

Statistic	FW.Br	FT.Br
Number of messages	5,167	8,526
Mean (SD)	0.270 (0.551)	0.230 (0.570)
Median	0.365	0.302
Q1; Q3	−0.052; 0.746	−0.178; 0.731
Interquartile range	0.797	0.909
Negative (< -0.05)	25.4%	28.4%
Neutral	9.9%	10.8%
Positive (> 0.05)	64.8%	60.8%
Mann-Whitney $U = 22,820,276$, $p = 4.0 \times 10^{-4}$		
Cliff’s $\delta = 0.036$, 95% CI 0.016, 0.056 (negligible)		
$\chi^2 = 21.28$, $p = 2.4 \times 10^{-5}$, Cramér’s $V = 0.039$		

4.4 Topic Modeling

In this work, the experimental evaluation included topic modeling methods LDA, PTM, and BERTopic. Table 4 illustrates the number of topics on each method by dataset. These methods were evaluated across two dimensions: topic coherence and topic diversity. Coherence was assessed through the Coherence C_V [39] and Coherence NPMI [31] metrics. Topic diversity, in turn, was measured by the diversity [20] and Inverted Rank-Biased Overlap (IRBO) [8] metrics. For the latter, we used the implementation provided by the researcher Silvia Terragni⁷. Each method selected the number of topics that best represented the data.

⁶Taking the group rather than the message as the unit of analysis, the difference is no longer detectable ($p = 0.61$ for groups with at least 10 messages).

⁷<https://github.com/silviatti/topic-model-diversity>

To identify the ideal number of topics (k), the C_V coherence metric was employed, varying k from 2 to 30 while maintaining the default parameter values for each method. For BERTopic, we employed the multilingual model paraphrase-multilingual-MiniLM-L12-v2 to generate *embeddings*. Next, we applied UMAP [34] for dimensionality reduction, using the following parameters: $n_neighbors=15$, $n_components=5$, $metric='cosine'$, and $random_state=42$. In the clustering stage, we used HDBSCAN [11], with the parameters $min_cluster_size = 30$, $metric='euclidean'$, and $prediction_data=True$. Table 5 presents the values obtained by LDA, PT Model, and BERTopic for the following metrics: Coherence NPMI, Coherence C_V, IRBO, and Diversity.

The results reported in Table 5 show that BERTopic achieved the best overall performance in terms of topic coherence for both datasets, obtaining the highest C_NPMI and C_V scores in FW.Br and FT.Br. For FW.Br, BERTopic reached a C_NPMI of 0.231 and a C_V of 0.659, outperforming both LDA and the PT Model. Similarly, for FT.Br, BERTopic obtained the highest coherence values, with a C_NPMI of 0.155 and a C_V of 0.641. Regarding topic distinctiveness, BERTopic also achieved the highest IRBO scores in both datasets, suggesting lower redundancy among the ranked topic terms. However, the PT Model produced the highest diversity score for FW.Br, while LDA achieved the highest diversity score for FT.Br, indicating that the most coherent models were not necessarily those with the greatest lexical diversity. Overall, these findings suggest that BERTopic was the most robust method for capturing semantically coherent topics across both platforms, whereas the diversity-oriented metrics varied according to the dataset and modeling strategy.

Table 4: Number of Topics on each Model by Dataset

Dataset	Amount of Topics		
	LDA	PTModel	BERTopic
FW.Br	6	2	5
FT.Br	2	28	26

We conducted a qualitative analysis to infer the main themes addressed and the latent social dynamics present in the two investigated datasets. The most relevant topics and their most representative terms are presented in Tables 6 and 7, for the FW.Br and TW.Br datasets, respectively.

For the FW.Br dataset, the LDA model identified six topics. Overall, all topics exhibit a strong presence of religious discourse, often intertwined with nationalist terminology. The first topic appears to group messages that employ Bolsonaro-aligned slogans, with emphasis on the terms God, all, above, Brazil, and truth. The second topic shares certain similarities with the first; however, it accentuates the religious bias, characterized by the words: “peace”, “pray”, and “word”. The third topic displays a markedly religious focus, highlighting terms such as glory, holy, victory, mighty, honor, wonderful, praise, and glorious.

Table 5: Topic Quality Evaluation Metrics

Model	LDA				PT Model				BERTopic			
	C_NPMI	C_V	IRBO	DIVERSITY	C_NPMI	C_V	IRBO	DIVERSITY	C_NPMI	C_V	IRBO	DIVERSITY
FW.Br	0,002	0,530	0,41	0,48	0,061	0,610	0,70	0,95	0,231	0,659	0,98	0,84
FT.Br	-0,008	0,514	0,58	0,80	0,024	0,563	0,94	0,73	0,155	0,641	0,99	0,78

Using PT Model, two main topics were identified for the FW.Br dataset. The first has a clearly religious focus, centered on traditional Christian discourse, with emphasis on terms such as “amen”, “prayer”, and “Father”. The second topic suggests messages associated with Bolsonaro-aligned slogans, including the words “God”, “above”, “all”, and “Brazil”.

BERTopic identified five main topics in the FW.Br dataset. The first topic exhibits strong religious content centered on Christianity, including terms such as “Jesus”, “life”, and “love”, suggesting a devotional tone commonly found in messages widely shared on social media. The second topic reflects content closely aligned with the Bolsonaro movement, as indicated by terms such as “God”, “above”, “all”, “Brazil”, “people”, and “nation”. The third topic is characterized by a more urgent and emotional tone, including terms related with sensationalist content, including “scenes”, “photos”, “violence”, and “share/forward”. This topic also contains references to Ricardo Nunes, then a candidate for mayor of São Paulo, through terms such as “undecided” and “Nunes”.

For the FT.Br dataset, the LDA model identified two main topics. The first highlights devotional communication characterized by expressions of gratitude and requests for blessings. Its vocabulary suggests personal messages of faith, thanksgiving, and protection, often associated with everyday experiences or immediate circumstances. The second topic further emphasizes the centrality of divine figures, with terms such as “God”, “Jesus”, “Christ”, and “Lord”. At the same time, it adopts a more assertive and collective tone, as indicated by expressions such as with “me”, and “repeat”, which suggest an effort to engage or influence the interlocutor.

PT Model identified 28 distinct topics on FT.Br dataset. Overall, most of these topics exhibit a strong presence of religious content, emphasizing evangelism, praise, and prayer. The first topic is characterized by terms such as “general”, “command”, “military”, and “army”, suggesting messages associated with the Armed Forces and patriotic-militaristic rhetoric. The second topic highlights expressions related to Bible reading and calls for personal conversion (“Bible”, “read”, and “written”). The third topic aggregates elements related to conservative values and ideological identity.

Finally, applying BERTopic to the FT.Br dataset produced 26 topics. The first one contains terms strongly associated with Christian discourse, reflecting themes such as faith, prayer, and devotion. The second topic is marked by the presence of proper names, which suggests multimedia content, such as the popular recording of Derci Gonçalves praising Jesus Christ. The third topic, in turn, addresses international themes, with emphasis on economic and public health issues.

4.5 Linguistic Analysis

To analyze the emotional, cognitive, and structural aspects of the texts in the investigated datasets, the LIWC tool was employed. LIWC extracts psycholinguistic features from textual data based on a predefined lexicon. In this study, we used the dictionary developed for Brazilian Portuguese, LIWC 2007 Português [7], which contains 127,149 words associated with 64 linguistic and psychological categories. These categories represent different linguistic and psychological dimensions, allowing the identification of patterns related to aspects such as emotions, cognitive processes, and linguistic structure.

For each dataset, the number of words associated with each LIWC category was counted after the text preprocessing step. The proportion of words belonging to each category was then calculated relative to the total number of words in the corresponding dataset. Thus, for each LIWC category, the analysis considered the proportion of words associated with that category and the proportion of words not associated with it. The resulting proportions were subsequently compared between the *FT.Br* and *FW.Br* datasets.

For each category, the difference between the proportions observed in the two datasets was calculated as $p_{Telegram} - p_{WhatsApp}$. A positive value therefore indicates a higher proportion of words associated with the category in *FT.Br*, whereas a negative value indicates a higher proportion in *FW.Br*. To assess whether the observed differences in proportions were statistically significant, a *chi-square test of independence* was performed independently for each LIWC category with a significance level of $\alpha = 0.05$.

From the 63 categories present in both datasets, 49 differ significantly at $\alpha = 0.05$ (40 after Holm-Bonferroni correction). However, all effect sizes are negligible ($\phi \leq 0.033$, median 0.008): significance here follows from the size of the token samples rather than from substantive divergence. Interpreting the differences in magnitude rather than in significance, the categories with largest differences are *time* (6.87% in FW.Br vs 8.77% in FT.Br), *relig* (12.20% vs 10.11%), *humans* (13.92% vs 11.85%), *money* (2.50% vs 3.38%) and *we* (0.74% vs 0.41%). The pattern seems coherent since FW.Br favours a devotional and communal register, with more religious vocabulary, more references to people and roughly twice the proportion of first-person plural pronouns, whereas FT.Br favours a more expository and transactional register, with more temporal, financial and causal vocabulary. Notably, the categories most directly related to hostility do not differ between the platforms (*anger*, $p = 0.41$; *anx*, $p = 0.46$), which is consistent with the polarity results reported in Section 4.3.

Table 6: Topics Generated by the FW.Br Dataset

Topics	LDA	PT Model	BERTopic
1	God; everyone; let us; Jesus; Brazil; above; Lord; life; love; because; bless; amen; mercy; good; always; faith; Father; are; about; time; do; nothing; truth; name; country; nation; can; today; also; people	God; Lord; Jesus; name; life; Christ; because; amen; prayer; Father	God; Jesus; Lord; everyone; name; life; let us; because; love; Christ
2	God; Lord; Jesus; Christ; name; let us; Father; days; iFood; nation; people; life; also; prayer; faith; pay; everyone; peace; right; pray; amen; because; about; love; word; happen; mercy; thus; can	God; everyone; let us; Brazil; faith; above; love; do; also; bless	Brazil; God; above; everyone; let us; people; Lord; nation; Jesus; Brazilian
3	God; Lord; Jesus; everyone; life; people; Brazil; name; glory; prayer; Christ; amen; love; because; holy; victory; children; powerful; days; honor; have; power; favor; do; things; are; wonderful; praise; dawn; glorious	—	scenes; photos; best; 100; undecided; violence; latest; share; true; Nunes

Table 7: Topics Generated by the FT.Br Dataset

Topics	LDA	PT Model	BERTopic
1	God; thank you; bless; today; never; Jesus; everyone; because; life; Brazil; nothing; me; reach; still; good; new; have; people; yesterday; future; truth; about; goals; are; book; better; ahead; congratulations; profit	general; command; military; BSB; goodbye; chief; RJ; army; bye; dear	God; Jesus; Lord; everyone; people; Christ; name; let us; truth; because
2	God; everyone; Jesus; with me; exists; repeat; Brazil; Lord; above; let us; now; nothing; Christ; name; can; country; power; nation; channel; kingdom; each; bless; prayer; Father; no one; mercy; love	Jesus; Bible; never; God; Christ; saves; forget; read; Mary	Derci; love; video; above; Ramiro; colonel; ITA; Jesus; perfect
3	—	Jews; conservative; religion; truth; want; say; account; any; such; conservatives	vaccine; covid19; vaccines; United; vaccinated; kingdom; study; Moderna; individuals

5 CONCLUSION AND FUTURE WORK

In this work, we analyze fundamentalist religious content disseminated in public WhatsApp and Telegram groups associated with the Brazilian far right. To do this, we provide two datasets and apply different data analysis techniques. It is essential to highlight that data from WhatsApp and Telegram, particularly from genuine groups, are scarce yet highly valuable to the scientific community, as they provide unique insights into contemporary Brazilian social dynamics, given the widespread adoption of these platforms in the country.

The results of the analyses conducted indicated that Brazilian fundamentalist discourse on WhatsApp and Telegram is multifaceted, combining religious, political, emotional, and ideological elements. Different algorithms capture these facets differently. This study showed that religious fundamentalist contents are present in far-right public groups (RQ1). The two platforms seems quite equivalent in emotional polarity since messages on Telegram are only marginally more negative than those on WhatsApp, with a negligible effect size (Cliff's $\delta = 0.036$), and both corpora are predominantly positive in lexical valence (RQ2). Topic analysis revealed distinct topics for each examined platform. Furthermore, the best-performing algorithm (in terms of coherence and diversity) also varied according to the platform (RQ3). The LIWC-based linguistic analysis identified significant differences in 49 of the 63 psycholinguistic categories, but with uniformly negligible effect sizes (RQ4).

Among the limitations of the present research, we highlight the following aspects. The extraction of religious messages was based on keywords provided by experts, which may introduce selection bias. Additionally, we deliberately focused on public groups associated with far-right discourse, as our objective was not to generalize religious discourse in Brazil, but rather to characterize a specific sociotechnical ecosystem. So, this choice may bias the analyses performed. As part of future work, we intend to conduct a temporal analysis investigating the variation of the themes addressed over time. A more detailed discourse analysis of the messages can reveal more subtle nuances in the language used to express fundamentalist views. Ultimately, we aim to investigate whether the results obtained can be generalized to other countries.

ETHICAL CONSIDERATIONS

This study was conducted exclusively with messages collected from public WhatsApp and Telegram groups, without recruitment, intervention, or direct interaction with the individuals whose messages were analyzed. The research follows the applicable provisions of Brazilian National Health Council Resolution CNS No. 510/2016 concerning studies based on publicly accessible information, as well as the ethical principles established by Resolutions CNS No. 466/2012 and No. 674/2022.

Because the study involved the processing of personal data and potentially sensitive information, particularly content

related to religious beliefs and political opinions, data handling procedures were also designed in accordance with the Brazilian General Data Protection Law (LGPD, Law No. 13,709/2018). Identifying attributes available in the original datasets, such as telephone numbers and group identifiers, were pseudonymized using cryptographic hash functions before the analytical stages. These identifiers were not used to identify individual users, and the analyses focused on aggregate, thematic, discursive, and psycholinguistic patterns rather than on profiling specific individuals.

To minimize the risk of re-identification, identifying information concerning non-public individuals is not intentionally reported in the manuscript, and only information necessary to address the research objectives was retained in the analyses. No images of children, adolescents, or other vulnerable individuals are presented. Data processing followed the principles of purpose limitation, necessity, security, and prevention established by the LGPD.

STATEMENT ON THE USE OF AI

Generative AI tools were used to assist with text structuring and language revision. Responsibility for the technical content, experiments, and verification of bibliographic references remains entirely with the authors.

ACKNOWLEDGMENTS

The authors would like to thank Lenovo for supporting the development of this research and the National Institute of Science and Technology for Combating Prejudice (INCT Preconceitos) for its institutional and scientific support.

REFERENCES

- [1] Imene Ajala, Shanaz Feroze, May El Barachi, Farhad Oroumchian, Sujith Mathew, Rand Yasin, and Saad Lutfi. 2022. Combining artificial intelligence and expert content analysis to explore radical views on twitter: Case study on far-right discourse. *Journal of Cleaner Production* 362 (2022), 132263. doi:10.1016/j.jclepro.2022.132263
- [2] Rafael J. A. Almeida. 2018. LeIA — Léxico para Inferência Adaptada. <https://github.com/rafjaa/LeIA>.
- [3] Gabriel A. Almond, R. Scott Appleby, and Emmanuel Sivan. 2003. *Strong Religion: The Rise of Fundamentalisms around the World*. University of Chicago Press, Chicago.
- [4] Victor Araújo. 2025. Does the growth of religious minorities transform electoral politics? Evidence from the evangelical boom in Brazil. *Political Science Research and Methods* (2025), 1–25. doi:10.1017/psrm.2025.10025
- [5] Karen Armstrong. 2000. *The Battle for God: Fundamentalism in Judaism, Christianity and Islam*. Knopf.
- [6] Andrew Atherstone and David Ceri Jones. 2023. *The Oxford Handbook of Christian Fundamentalism*. Oxford University Press. doi:10.1093/oxfordhb/9780198844594.001.0001
- [7] Pedro Balage Filho, Thiago Alexandre Salgueiro Pardo, and Sandra Aluísio. 2013. An evaluation of the Brazilian Portuguese LIWC dictionary for sentiment analysis. In *Proceedings of the 9th Brazilian Symposium in Information and Human Language Technology*.
- [8] Federico Bianchi, Silvia Terragni, and Dirk Hovy. 2021. Pre-training is a Hot Topic: Contextualized Document Embeddings Improve Topic Coherence. In *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing*. 759–766.
- [9] David M. Blei, Andrew Y. Ng, and Michael I. Jordan. 2003. Latent Dirichlet Allocation. *Journal of Machine Learning Research* 3, null (mar 2003), 993–1022.
- [10] Carlos Arcila Calderón, Gonzalo de la Vega, and David Blanco Herrero. 2020. Topic Modeling and Characterization of Hate Speech against Immigrants on Twitter around the Emergence of a Far-Right Party in Spain. *Social Sciences* 9, 11 (2020). doi:10.3390/socsci9110188
- [11] Ricardo J. G. B. Campello, Davoud Moulavi, Arthur Zimek, and Jörg Sander. 2015. Hierarchical Density Estimates for Data Clustering, Visualization, and Outlier Detection. *ACM Transactions on Knowledge Discovery from Data* 10, 1, Article 5 (2015), 51 pages. doi:10.1145/2733381
- [12] Brenda Carranza and Maria José F. Rosado-Nunes. 2024. Democracy and the Christian Right in Brazil: Family, Sexualities and Religious Freedom. *Religions* (2024). <https://api.semanticscholar.org/CorpusID:270021489>
- [13] Flavio Carvalho, Rafael Rodrigues, Gabriel Santos, Pedro Cruz, Lilian Ferrari, and Gustavo Guedes. 2019. Avaliação da versão em português do LIWC Lexicon 2015 com análise de sentimentos em redes sociais. In *Anais do VIII Brazilian Workshop on Social Network Analysis and Mining* (Belém). SBC, Porto Alegre, RS, Brasil, 24–34. doi:10.5753/brasnam.2019.6545
- [14] Guilherme Casarões and Amy Erica Smith. 2023. “Brazil above Everything or God above Everyone?”: The Sources of Brazil’s Religious Soft Power. In *The Geopolitics of Religious Soft Power: How States Use Religion in Foreign Policy*. Oxford University Press. arXiv:<https://academic.oup.com/book/0/chapter/413455769/chapter-pdf/58132395/oso-9780197605806-chapter-14.pdf> doi:10.1093/oso/9780197605806.003.0014
- [15] Athus Cavallini, Thamya Donadia, and Giovanni Comarela. 2024. Characterizing the Toxicity of the Brazilian Extremist Communities on Telegram. In *Proceedings of the 30th Brazilian Symposium on Multimedia and the Web* (Juiz de Fora/MG). SBC, Porto Alegre, RS, Brasil, 370–374. doi:10.5753/webmedia.2024.243207
- [16] Alexis Conneau, Kartikay Khandelwal, Naman Goyal, Vishrav Chaudhary, Guillaume Wenzek, Francisco Guzmán, Edouard Grave, Myle Ott, Luke Zettlemoyer, and Veselin Stoyanov. 2020. Unsupervised Cross-lingual Representation Learning at Scale. *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics (ACL)* (2020).
- [17] Benjamin A. Cowan. 2021. *Moral Majorities Across the Americas: Brazil, the United States, and the Creation of the Religious Right*. University of North Carolina Press, Chapel Hill, NC. 304 pages.
- [18] Ivandro Claudino de Sá, Thiago Gadelha, Tiago Vinuto, José Wellington Franco da Silva, José Maria Monteiro, and Javam C. Machado. 2023. A Real-Time Platform to Monitoring Misinformation on Telegram. In *Proceedings of the ICEIS 2023*. 271–278.
- [19] Ivandro Claudino de Sá, Luciano Galic, Wellington Franco, Thiago Gadelha, José Maria Monteiro, and Javam C. Machado. 2023. BATMAN: A Big Data Platform for Misinformation Monitoring. In *Proceedings of the ICEIS 2023*. 237–246.
- [20] Adji B. Dieng, Francisco J. R. Ruiz, and David M. Blei. 2020. Topic Modeling in Embedding Spaces. *Transactions of the Association for Computational Linguistics* 8 (2020), 439–453.
- [21] Explosion AI. 2026. pt_core_news_sm: Portuguese pipeline optimized for CPU. <https://spacy.io/models/pt>. Accessed: Janeiro 2026.
- [22] Paula Gibrim, Marcus Ribeiro, Lorena Moreira, Philippe Melo, and Julio Reis. 2025. I’m not Welcome Here – Uma Revisão de Escopo sobre Toxicidade e Discurso de Ódio em Jogos Digitais. In *Proceedings of the 31st Brazilian Symposium on Multimedia and the Web* (Rio de Janeiro/RJ). SBC, Porto Alegre, RS, Brasil, 646–660. doi:10.5753/webmedia.2025.15998
- [23] Jarod Govers, Philip Feldman, Aaron Dant, and Panos Patros. 2023. Down the Rabbit Hole: Detecting Online Extremism, Radicalisation, and Politicised Hate Speech. *ACM Comput. Surv.* 55, 14s, Article 319 (jul 2023), 35 pages. doi:10.1145/3583067
- [24] Maarten Grootendorst. 2022. BERTopic: Neural topic modeling with a class-based TF-IDF procedure. *arXiv preprint arXiv:2203.05794* (2022).
- [25] Samuel Guimarães, Gabriel Kakizaki, Philippe Melo, Márcio Silva, Fabricio Murai, Julio C. S. Reis, and Fabricio Benevenuto. 2023. Anatomy of Hate Speech Datasets: Composition Analysis and Cross-dataset Classification. In *Proceedings of the 34th ACM Conference on Hypertext and Social Media* (Rome, Italy) (HT ’23). Association for Computing Machinery, New York, NY, USA, Article 33, 11 pages. doi:10.1145/3603163.3609158

- [26] Sami Abdullah Hamdi. 2022. Mining ideological discourse on Twitter: The case of extremism in Arabic. *Discourse & Communication* 16, 1 (2022), 76–92. arXiv:https://doi.org/10.1177/17504813211043706 doi:10.1177/17504813211043706
- [27] Elina Inkeri Hartikainen. 2025. Religious Intolerance and the Semiotics of Secular Indeterminacy in Brazil. *Signs and Society* 13, 1 (2025), 78–101. doi:10.1017/sas.2024.4
- [28] C. J. Hutto and Eric Gilbert. 2014. VADER: A Parsimonious Rule-based Model for Sentiment Analysis of Social Media Text. In *Proceedings of the Eighth International AAAI Conference on Weblogs and Social Media (ICWSM-14)*. 216–225.
- [29] Mi Kim and Dosung Kim. 2022. A Suggestion on the LDA-Based Topic Modeling Technique Based on ElasticSearch for Indexing Academic Research Results. *Applied Sciences* 12, 6 (2022). doi:10.3390/app12063118
- [30] Ugur Kursuncu, Manas Gaur, Carlos Castillo, Amanuel Alambo, Krishnaprasad Thirunarayan, Valerie Shalin, Dilshod Achilov, I. Budak Arpinar, and Amit Sheth. 2019. Modeling Islamist Extremist Communications on Social Media using Contextual Dimensions: Religion, Ideology, and Hate. *Proc. ACM Hum.-Comput. Interact.* 3, CSCW, Article 151 (nov 2019), 22 pages. doi:10.1145/3359253
- [31] Jey Han Lau, David Newman, and Timothy Baldwin. 2014. Machine Reading Tea Leaves: Automatically Evaluating Topic Coherence and Topic Model Quality. In *Proceedings of the 14th Conference of the European Chapter of the Association for Computational Linguistics*. 530–539.
- [32] João Lobo, Daniel Barbosa, and Philippe Melo. 2025. Investigando a Dinâmica da Propagação do Ódio em Cascatas de Toxicidade nos Chats Ao Vivo da Twitch. In *Proceedings of the 31st Brazilian Symposium on Multimedia and the Web (Rio de Janeiro/RJ)*. SBC, Porto Alegre, RS, Brasil, 321–329. doi:10.5753/webmedia.2025.15110
- [33] Victor Martins, Sophia Serafim, Lívia Pereira, Amanda Alves, Carlos Ferreira, and Jussara Almeida. 2025. A Misoginia no YouTube Brasileiro: Um Estudo de Caso sobre o Conteúdo Produzido pela Comunidade Red Pill. In *Proceedings of the 31st Brazilian Symposium on Multimedia and the Web (Rio de Janeiro/RJ)*. SBC, Porto Alegre, RS, Brasil, 28–37. doi:10.5753/webmedia.2025.16070
- [34] Leland McInnes, John Healy, Nathaniel Saul, and Lukas Großberger. 2018. UMAP: Uniform Manifold Approximation and Projection. *Journal of Open Source Software* 3, 29 (2018), 861. doi:10.21105/joss.00861
- [35] Cas Mudde. 2019. *The Far Right Today*. Polity Press, Cambridge, UK.
- [36] Fernanda Nascimento, Melissa Sousa, Gustavo Martins, José Maria Monteiro, and Javam C. Machado. 2025. Detecting Misinformation Virality on WhatsApp. In *Proceedings of the 14th International Conference on Data Science, Technology and Applications, DATA 2025, Bilbao, Spain, June 10-12, 2025*, Elhadj Benkhelifa, Xin-She Yang, and Slimane Hammoudi (Eds.). SCITEPRESS, 685–692. doi:10.5220/0013646800003967
- [37] Giovana Piorino, Luiz Lima, Adriana Pagano, and Ana Silva. 2025. Toxicidade e Gatilhos: Um Estudo de Caso em Comunidades do Reddit no Brasil. In *Proceedings of the 31st Brazilian Symposium on Multimedia and the Web (Rio de Janeiro/RJ)*. SBC, Porto Alegre, RS, Brasil, 575–579. doi:10.5753/webmedia.2025.15984
- [38] Andrea L. P. Pirro. 2022. Far right: The significance of an umbrella concept. *Nations and Nationalism* 29, 1 (2022), 101–112. doi:10.1111/nana.12860
- [39] Michael Röder, Andreas Both, and Alexander Hinneburg. 2015. Exploring the Space of Topic Coherence Measures. In *Proceedings of the Eighth ACM International Conference on Web Search and Data Mining, WSDM 2015, Shanghai, China, February 2-6, 2015*, Xueqi Cheng, Hang Li, Evgeniy Gabrilovich, and Jie Tang (Eds.). ACM, 399–408. doi:10.1145/2684822.2685324
- [40] Renato Russo, Ioana Literat, and Paulo Blikstein. 2023. Weaponizing John 8:32: The Complex Interconnections Between Religion and Politics on Brazilian Twitter. *Social Media + Society* 9, 4 (2023), 20563051231220328. arXiv:https://doi.org/10.1177/20563051231220328 doi:10.1177/20563051231220328
- [41] Yash Saini, Vishal Bachchas, Yogesh Kumar, and Sanjay Kumar. 2020. Abusive Text Examination Using Latent Dirichlet Allocation, Self Organizing Maps and K Means Clustering. In *2020 4th International Conference on Intelligent Computing and Control Systems (ICICCS)*. 1233–1238. doi:10.1109/ICICCS48265.2020.9121090
- [42] Silvia Terragni, Debasis Ganguly, and Fabio Crestani. 2020. OCTIS: Comparing and optimizing topic models is simple!. In *Proceedings of the 43rd International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR '20)*. ACM, 2201–2204. doi:10.1145/3397271.3401215
- [43] Javier Torregrosa, Gema Bello-Orgaz, Eugenio Martínez-Cámara, Javier Del Ser, and David Camacho. 2022. A survey on extremism analysis using natural language processing: definitions, literature review, trends and challenges: definitions, literature review, trends and challenges. *Journal of Ambient Intelligence and Humanized Computing* unknown, 8 (12 Dec. 2022), 9869–9905. doi:10.1007/s12652-021-03658-z Publisher Copyright: © 2022, The Author(s)..
- [44] Francielle Vargas, Isabelle Carvalho, Wolfgang Schmeisser-Nieto, Fabrício Benevenuto, and Thiago Pardo. 2023. NoHateBrazil: A Brazilian Portuguese Text Offensiveness Analysis System. In *Proceedings of the 14th International Conference on Recent Advances in Natural Language Processing*. INCOMA Ltd., Shoumen, Bulgaria, Varna, Bulgaria, 1180–1186. https://aclanthology.org/2023.ranlp-1.125
- [45] Yuan Zuo, Junjie Wu, Hui Zhang, Hao Lin, Fei Wang, Ke Xu, and Hui Xiong. 2016. Topic Modeling of Short Texts: A Pseudo-Document View. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (San Francisco, California, USA) (KDD '16)*. Association for Computing Machinery, New York, NY, USA, 2105–2114. doi:10.1145/2939672.2939880